Lecture 4: Predictive Method

Introduction

- The predictive method:
  - provides a structured methodology to estimate the expected average crash frequency (by total crashes, crash severity or collision type) of a site, facility or roadway network.
    - A site: a homogenous roadway segment or an individual intersection.
      - e.g., divided or undivided roadway segments, signalized or unsignalized intersections.
    - A facility: consists of a contiguous set of individual intersections and roadway segments.
      - e.g., rural two-lane two-way roads, multilane highways, urban and suburban arterials.
    - A road network: consists of a number of contiguous facilities.

Introduction

- The predictive method:
  - can be used for evaluating and comparing the expected average crash frequency of situations like:
    - Existing sites/facilities/networks under past or future traffic volumes;
    - Alternative designs for an existing site/facility/network under past or future traffic volumes;
    - Designs for a new site/facility/network under future (forecast) traffic volumes;
    - …

Introduction

- Procedures:
  - First, the predicted average crash frequency of an individual site ($N_{predicted}$) is estimated based on the geometric design, traffic control features, and traffic volumes of that site, using statistical models developed from data for a number of similar sites.
    - The estimate is for a given time period of interest during which the geometric design and traffic control features are unchanged and traffic volumes are known or forecast.
    - The estimate relies upon statistical models developed from data for a number of similar sites.

Introduction

- Procedures:
  - For an existing site, the observed crash frequency ($N_{observed}$) for that specific site is then combined with $N_{predicted}$ to improve the statistical reliability of the estimate.
  - The result from the predictive method is the expected average crash frequency ($N_{expected}$) of that site.
  - The cumulative sum of all sites is used as the estimate for an entire facility or network.

Introduction

- The predictive models vary by site and facility type but all have the same basic elements:
  - **Safety Performance Functions (SPFs)**: statistical “base” models that are used to estimate the predicted average crash frequency for specific site types with base conditions.
    - SPFs are typically a function of a number of variables, such as AADT.
  - **Crash Modification Factors (CMFs)**: the ratio of the effectiveness of one condition in comparison to another condition. CMFs are used to account for the difference between the specific site conditions and the base conditions.
    - e.g., the SPF for roadway segments has a base condition of 12-ft lane width, but the specific site may be a roadway segment with a 10-ft lane width.
  - **Calibration factor (C)**: is used to account for differences between the jurisdiction and time period.

Introduction

$$N_{predicted} =$$

$$SPF \times (CMF_1 \times CMF_2 \times \ldots.) \times C$$

where:

SPF = Safety Performance Function

CMF = Crash Modification Factor

C = Calibration Factor

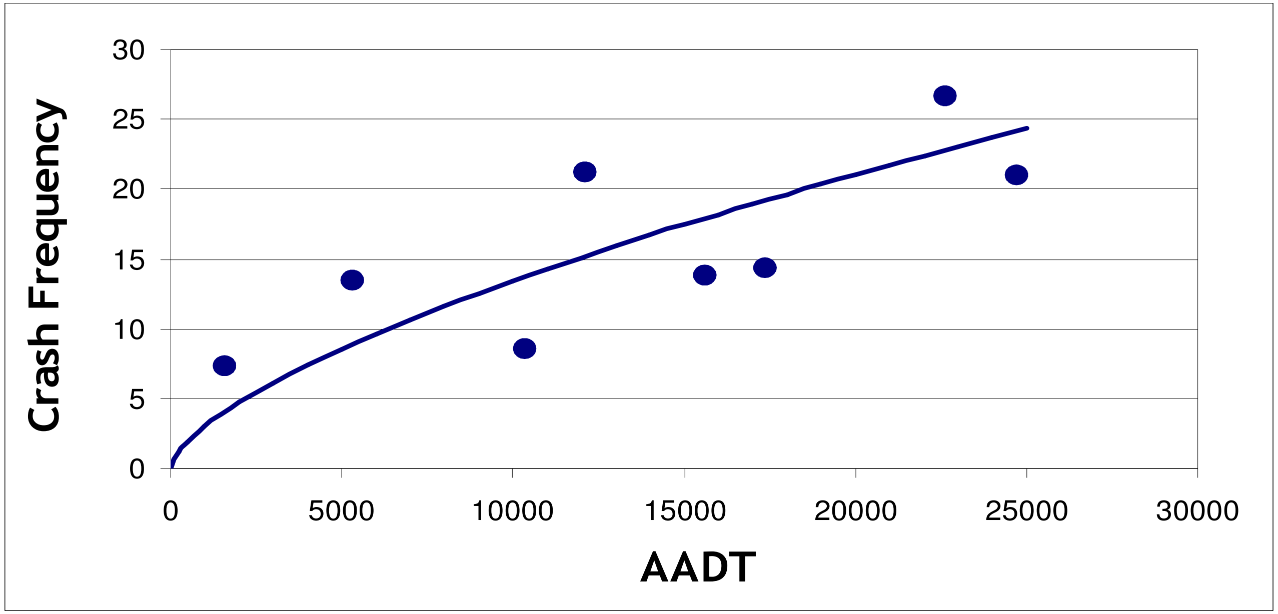
$$N_{expected} = N_{predicted} \oplus N_{observed}$$

Introduction

- Advantages
  - Regression-to-the-mean bias is addressed as the method concentrates on long-term expected average crash frequency rather than short-term observed crash frequency;
  - Reliance on availability of limited crash data for any one site is reduced by incorporating predictive relationships based on data from many similar sites;
  - The predictive method provides a method of crash estimation for sites or facilities that have not been constructed or have not been in operation long enough to make an estimate based on observed crash data;

Introduction

- Advantages (cont.)
  - The method accounts for the fundamentally nonlinear relationship between crash frequency and traffic volume.



Safety Performance Functions

- SPFs in HSM are
  - regression equations that estimate the average crash frequency for a specific site type (with specified base conditions) as a function of annual average daily traffic (AADT) and, in the case of roadway segments, the segment length (L).

Safety Performance Functions

- A SPF for roadway segments on rural two-lane highways:

$$N_{SPF\ rs} = (AADT) \times (L) \times (365) \times 10^{(-6)} \times e^{(-0.4865)}$$

where
AADT = annual average daily traffic volume (v/d)
on road way segment.
L: length of roadway segment (miles)

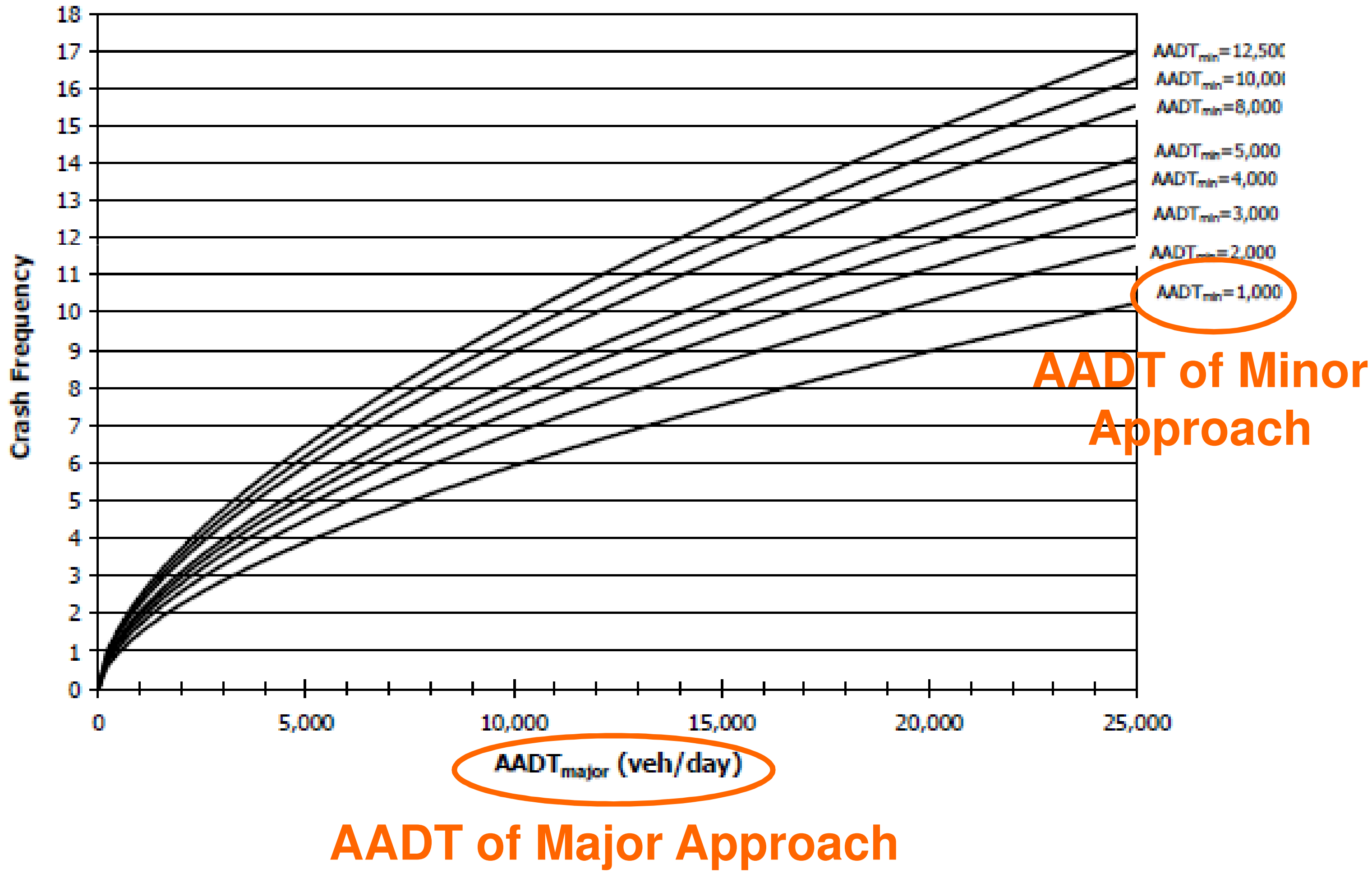
- A SPF for four-leg signalized intersections:

$$N_{spf\ 4SG} = exp[-5.13 + 0.60 \times \ln(AADT_{maj}) + 0.20 \times \ln(AADT_{min})]$$

where
$AADT_{maj}$ = AADT (v/d) on the major road.
$AADT_{min}$ = AADT (v/d) on the minor road.

Safety Performance Functions

The SPF for Four-leg Signalized Intersections



Crash Prediction Models

- Multivariate models
- Fitted to crash data
- Statistical relationship between the number of crashes and factors causally related to crashes

Crash Prediction Models

- Choice of explanatory variables
  - Crash rate (or risk) is traditionally defined as the number of crashes per unit of exposure

$$Crash\ Rate = \frac{Average\ Crash\ Frequency\ in\ a\ Period}{Exposure\ in\ Same\ Period}$$

  - Expected number of crashes = Exposure $\times$ Risk
  - Therefore, the explanatory variables:
    - Variables describing exposure
    - Risk factors

Crash Prediction Models

- Choice of explanatory variables
  - The usual basis for choosing explanatory variables appears to be simply **data availability.**
  - Explanatory variables should:
    - have been found in previous studies to have a major influence on the number of crashes;
    - can be measured in a valid and reliable way;
    - are not very highly correlated with other explanatory variables included.

Crash Prediction Models

- Variables that are usually included:
  - exposure variable (e.g., vehicle kilometers traveled)
  - variables describing the transport function of the road (motorway, main arterial, …)
  - variables describing cross section (number of lanes, lane width, shoulder width, …)
  - variables describing traffic control (speed limit, type of traffic control at intersections, …)
- Variables that are less often included:
  - variables describing alignment
  - estimates of pedestrian and cyclist exposure
  - variables describing road user behavior

Crash Prediction Models

- Choice of model form
  - The basic form of nearly all modern crash prediction models:

$$E(\lambda) = \alpha Q^{\beta} e^{\sum \gamma_i x_i}$$

  - A power function is applied for exposure, and
  - An exponential function applied for risk factors.
  - Additive, linear models are rarely used. Why?
    - e.g., negative predicted number of crashes.

Crash Prediction Models

- Generalized linear models (GLM)
  - A flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution.
- The GLM consists of three components:
  - A random component
    - specifying the conditional distribution of the response variable (e.g., Poisson, binomial, gamma, …)
  - A linear predictor
    - a linear function of regressors

$$\eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

Crash Prediction Models

- The GLM consists of three components (cont.):
  - An invertible link function $g(\cdot)$
    - transforming the expectation of the response variable ($\mu_i \equiv E(Y_i)$)

$$g(\mu_i) = \eta_i$$

Link	$\eta_i = g(\mu_i)$	$\mu_i = g^{-1}(\eta_i)$
Identity	μ_i	η_i
Log	$\log_e \mu_i$	e^{η_i}
Inverse	μ_i^{-1}	η_i^{-1}
Inverse-square	μ_i^{-2}	$\eta_i^{-1/2}$
Square-root	$\sqrt{\mu_i}$	η_i^2
Logit	$\log_e \frac{\mu_i}{1-\mu_i}$	$\frac{1}{1+e^{-\eta_i}}$

Crash Prediction Models

- Evaluation of goodness of fit
  - Log-likelihood ratio
  - Pearson's chi-squared test
  - Mean deviance ratio
  - Akaike’s Information Criterion (AIC)
  - The Freeman-Tukey index
  - t-Statistic
  - Elvik index

Crash Prediction Models

- Total variation = Random variation + Systematic variation
- Only the systematic part can be explained by means of crash prediction models
- Systematic variation is caused whenever the variance > the mean
  - This is referred to as overdispersion

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\lambda = \frac{\frac{\mathrm{Var}(x)}{\mu} - 1}{\mu}$$

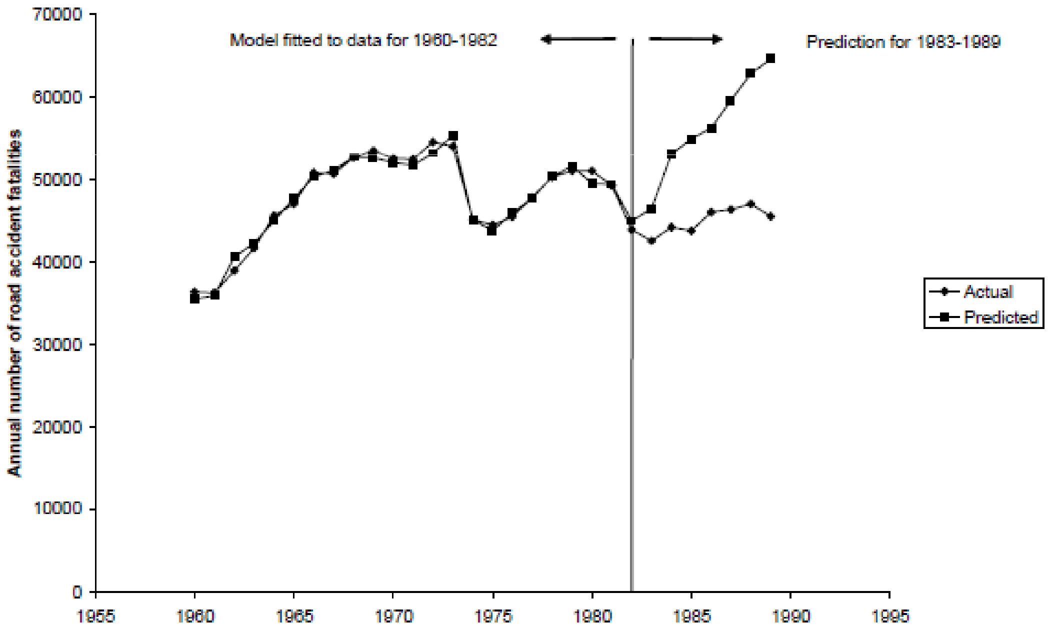
λ: overdispersion parameter
μ: mean

Crash Prediction Models

	Distribution of road sections by number of fatalities			Accident prediction model	
Number of fatalities	Actual	Poisson	Negative binomial	Explanatory variables	Coeff.
0	19,957	19,728	19,974	Constant	-7.154
1	895	1274	854	ln(AADT)	0.842
2	135	41	163	Speed limit 50 km/h	Reference category
3	43	1	39	Speed limit 60 km/h	-0.020
4	9	0	10	Speed limit 70 km/h	0.385
5	3	0	3	Speed limit 80 km/h	0.172
6	1	0	1	Speed limit 90 km/h, rural road	0.090
7	0	0	0	Speed limit 90 km/h, class B road	0.610
8	1	0	0	Speed limit 90 km/h, class A road	0.879
N	21,044	21,044	21,044	Number of lanes	-1.967
				Number of intersections/km	0.082
				Dummy for trunk road	0.255
Mean	0.0646				
Variance	0.0976			Estimated variance	0.0745
Overdispersion parameter	7.91			Overdispersion parameter	2.39

Crash Prediction Models

- Predictive performance assessment
  - predictive models ≠ explanatory models



Perfectly reproducing the data it was fitted to, yet giving badly wrong predictions for future years.

Crash Prediction Models

- Predictive performance assessment
  - Using the model to predict crash counts in future years
  - Splitting data into "training set" and "test set"
- Cross-validation technique
  - Exhaustive cross-validation
    - Leave-one-out cross-validation
    - Leave-p-out cross-validation
  - Non-exhaustive cross-validation
    - 2-fold cross-validation
    - k-fold cross-validation
    - Repeated random sub-sampling validation

Crash Prediction Models

- Potential sources of errors
  - Omitted variable bias (mostly exposure variables)

$$\text{Number of pedestrian accidents} = 0.0000734 \times MV^{0.50} \times PED^{0.72}$$

- MV from 5,000 to 10,000
- PED from 500 to 1000

→ No. of pedestrian crashes?

- PED from 100 to 1000
- PED from 1000 to 2000

→ Injury rate per pedestrian?

- What if PED was not included in the model?
  - MV's exponent would change from 0.5 to 0.9
  - MV's coefficient contains part of the effect of pedestrian volume when that is not included

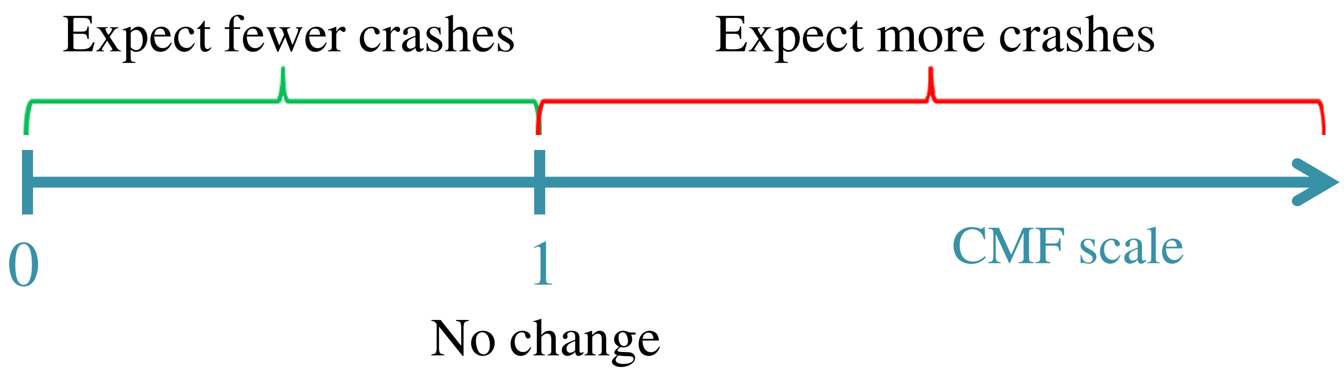
Crash Prediction Models

- Potential sources of errors (cont.)
  - (Multi)co-linearity among explanatory variables
    - Lead to unstable estimates of the coefficients
    - Variance Inflation Factor (VIF) test
  - Wrong functional form for relationships between variables
    - Occurs when a single function is used while relationship varies, depending on circumstances, e.g., day-time vs. night-time crashes.
    - Occurs when traffic volume is represented by an average value, like AADT rather than actual data, e.g., traffic volume varies during the day and from day to day.

Crash Modification Factors

- A CMF is an index that quantifies the change in crash frequency at a site as a result of implementing a specific treatment or countermeasure.

$$CMF_i = \frac{\text{expected crash frequency if change } i \text{ is made}}{\text{expected crash frequency if change } i \text{ is not made}}$$



Crash Modification Factors

- CMF vs. CRF
  - Crash Reduction Factor (CRF) assumes a reduction in crashes due to implementing a countermeasure.
  - Example: if CMF for a treatment is 0.88, what is the corresponding CRF?

CMF = 0.88 => expected crashes after treatment is 88% of crashes before treatment.

CRF = 1.00 – CMF

i.e., 1.00 – 0.88 = 0.12, or 12% reduction

Crash Modification Factors

- CMFs are estimated based on statistical analyses of reported crash data
  - Before-after studies (with or without comparison group)
    - The same set of sites are used and the CMF is estimated by examining safety performance before the treatment is implemented and after the treatment is implemented.
    - A comparison group of sites is used to provide a baseline for how safety performance changes when the treatment is not applied.
  - Cross-sectional studies (with or without regression)
    - Identify sites both with and without treatment in the same time period to compare how the treatment impacts safety performance.
    - Regression is used to help control for the impacts of other factors that might have.

Crash Modification Factors

- Each CMF value only applies to a very specific set of conditions.
  - Area type: urban, suburban, rural
  - Crash type: all, run-off-road, night, multi-vehicle, etc.
  - Crash severity: fatality, serious injury, slight injury, PDO
  - Roadway volumes: typically measured in AADT
  - Roadway geometry: number of lanes, number of legs at an intersection, etc.
  - Traffic control: speed limit, type of intersection control, etc.
  - …

Crash Modification Factors

- Example:
  - Improperly defined: CMF for edgeline rumble strips
  - Properly defined: CMF for edgeline rumble strips on fatal run-off-the-road crashes on two-lane rural roads
- Reasons for set of conditions
  - Specific countermeasures/treatments only impact a specific subset of crash types.
  - Same countermeasure/treatment in different contexts or driving environments may have different effects.
  - CMFs are sometimes estimated with only a certain type of reported crash data.

Crash Modification Factors

- CMF applications
  - Infrastructure treatments
    - Widening lanes or shoulders
    - Install rumble strips
  - Traffic control
    - Signing, pavement markings
    - Signalization
  - Operational strategies
    - Access management (e.g., driveway closure, median closures)
  - Maintenance strategies
    - Anti-icing applications
  - Enforcement strategies
    - Automated enforcement

CMF
CRASH MODIFICATION FACTORS CLEARINGHOUSE

Skip to main content | Site Map | Notice | Home

About CMFs | Find CMFs | Submit CMFs | Resources | Contact

Quick Search

enter search term(s)

- narrow by countermeasure category -
- narrow by crash type -
- narrow by crash severity -
- narrow by roadway type -

Advanced Search

Search CMFs

Need Help?

A crash modification factor (CMF) is a multiplicative factor used to compute the expected number of crashes after implementing a given countermeasure at a specific site. The Crash Modification Factors Clearinghouse houses a Web-based database of CMFs along with supporting documentation to help

Recently Added CMFs

Design diamond, trumpet or cloverleaf interchange	Physical channelization of left-turn lane on major road	Flashing beacons at four leg stop controlled intersections
CMF: 0.96	CMF: 0.73	CMF: 0.87
CRF: 4	CRF: 27	CRF: 13

www.cmfclearinghouse.org

CMF	CRF(%)	Quality	Crash Type	Crash Severity	Roadway Type	Area Type	Reference
0.71 [B]	29	★★★★★	All	Serious injury,Minor injury	Minor Arterial	Urban	Elvik, R. and Vaa, T., 2004

Category:Advanced technology and ITS

Countermeasure: Install red-light cameras at intersections

CMF	CRF(%)	Quality	Crash Type	Crash Severity	Roadway Type	Area Type	Reference
0.84 [B]	16	★★★★☆	Angle	Serious injury,Minor injury	Not specified	Urban	Persaud et al., 2005
1.24 [B]	-24	★★★★☆	Rear end	Serious injury,Minor injury	Not specified	Urban	Persaud et al., 2005

Category:Highway lighting

Countermeasure: Illumination

CMF	CRF(%)	Quality	Crash Type	Crash Severity	Roadway Type	Area Type	Reference
0.62 [B]	38	★★★★☆	All	Serious injury,Minor injury	Not specified	Not specified	Elvik, R. and Vaa, T., 2004

www.cmfclearinghouse.org

Crash Modification Factors

- Example use of CMFs
  - Two-lane rural highway segment
  - AADT 2008 = 4494 v/d
  - Current: 12’ lanes
  - Proposed: 11’ lanes

Year	Total segment crashes
2006	26
2007	19
2008	17
Total	62
Average	20.67

What will be the expected number of crashes after change?

Crash Modification Factors

- Example use of CMFs

Exhibit 13-2: AMF for Lane Width on Rural Two-Lane Roadway Segments[(16)]

Lane Width	Average Annual Daily Traffic (AADT) (vehicles/day)		
	< 400	400 to 2000	> 2000
9 ft or less	1.05	$1.05+2.81\times10^{-4}(AADT-400)$	1.50
10 ft	1.02	$1.02+1.75\times10^{-4}(AADT-400)$	1.30
11 ft	1.01	$1.01+2.5\times10^{-5}(AADT-400)$	1.05
12 ft or more	1.00	1.00	1.00

NOTE: The collision types related to lane width to which these AMFs apply are single-vehicle run-off the-road and multiple-vehicle head-on, opposite-direction sideswipe, and same-direction sideswipe accidents.

Standard error of the AMF is unknown.

Crash Modification Factors

- Example use of CMFs

Year	Total segment crashes	Run-off-the-road, head-on, and sideswipe crashes
2006	26	10
2007	19	13
2008	17	10
Total	62	33
Average	20.67	11.00

$$11.00 / 20.67 = 0.53 \text{ or } 53\%$$

Crash Modification Factors

- Example use of CMFs

CMF Conversion:

$$CMF = [(CMF_{ra} - 1.0) \times p_{ra}] + 1.0$$

where
CMF: Crash Modification Factor for total crashes
CMF_{ra}: Crash Modification Factor for related crashes
p_{ra}: proportion of total crashes constituted by related crashes

11' lane width with $CMF_{ra} = 1.05$
$p_{ra} = 0.53$

$$CMF = [(1.05 - 1.0) \times 0.53] + 1.0$$
$$= 1.0265$$

Crash Modification Factors

- CMFs can be multiplied together to estimate the combined effects of different countermeasures/ treatments that have independent effects.

$$N_{predicted} = N_{base\ condition} \times (CMF_1 \times CMF_2 \times \ldots.)$$

Example

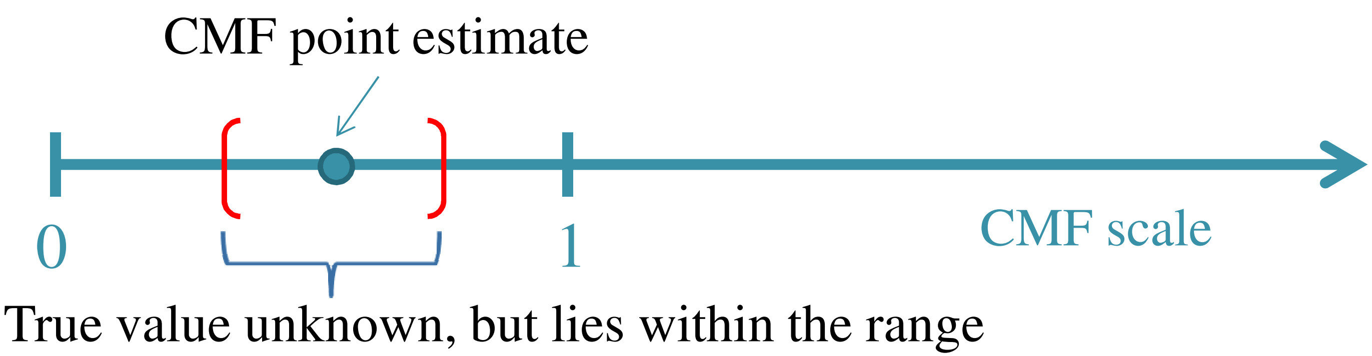
Treatment 'x' consists of providing a left-turn lane on both major-road approaches to an urban four-leg signalized intersection and treatment 'y' is permitting right-turn-on-red maneuvers. These treatments are to be implemented and it is assumed that their effects are independent of each other. An urban four-leg signalized intersection is expected to have 7.9 accidents/year. For treatment t_x, AMF_x = 0.81; for treatment t_y, AMF_y = 1.07.

What accident frequency is to be expected if treatment x and y are both implemented?

expected accidents = 7.9 x 0.81 x 1.07 = 6.8 accidents/year.

Crash Modification Factors

- Errors in CMFs
  - Errors may exist due to:
    - Type of statistical model
    - Amount of crash data
    - Variation in crash data
    - Crash data reporting
- Numerical value of a CMF is a point estimate
  - Subject to some amount of uncertainty



Crash Modification Factors

- Standard error of the CMF
  - Most studies not only provide the point estimate of the CMF but also provide an estimate of the amount of error associated with the point estimate.
- Standard error gives indication of precision
  - Small standard error → precise estimate
  - Large standard error → imprecise estimate

Crash Modification Factors

- Confidence interval (CI) for CMF
  - Combine point estimate and standard error to estimate the range that the true CMF value is believed to lie within.

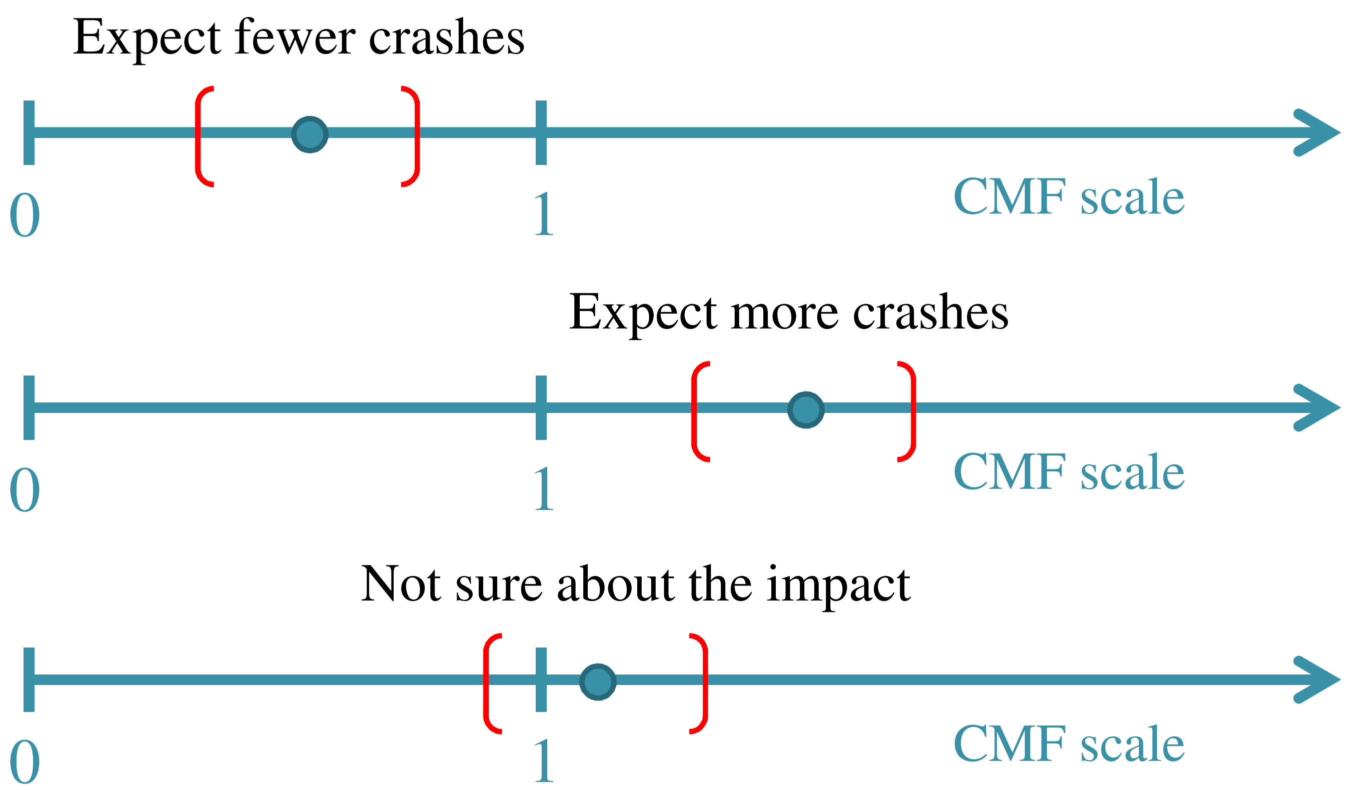
$$CI_{CMF} = P_{CMF} \pm z \times ERROR$$

where z is associated with the level of certainty or confidence that we would like to have.

Type of confidence interval	Z value
90% confidence interval	1.64
95% confidence interval	**1.96**
99% confidence interval	2.58

Crash Modification Factors

- Using CI provides a better indication of expected impacts of a countermeasure/treatment.



Calibration Factor

- The main purpose of the calibration procedure is to adjust predictive models that were developed with data from on jurisdiction for application in another jurisdiction.
- Calibration provides a method to account for difference between jurisdictions:
  - Climate
  - Driver populations
  - Animal populations
  - Crash reporting thresholds
  - Crash reporting system procedures

Calibration Factor

- The calibration procedure involves five steps:
  - Step 1 Identify facility types for which the predictive model is to be calibrated
  - Step 2 Select sites for calibration of the predictive model for each facility type
  - Step 3 Obtain data for each facility type applicable to a specific calibration period
  - Step 4 Apply the predictive model to predict total crash frequency for each site during the calibration period as a whole
  - Step 5 Compute calibration factors

Calibration Factor

- Example:
  - The SPF for four-leg signalized intersections on rural two-lane roads is

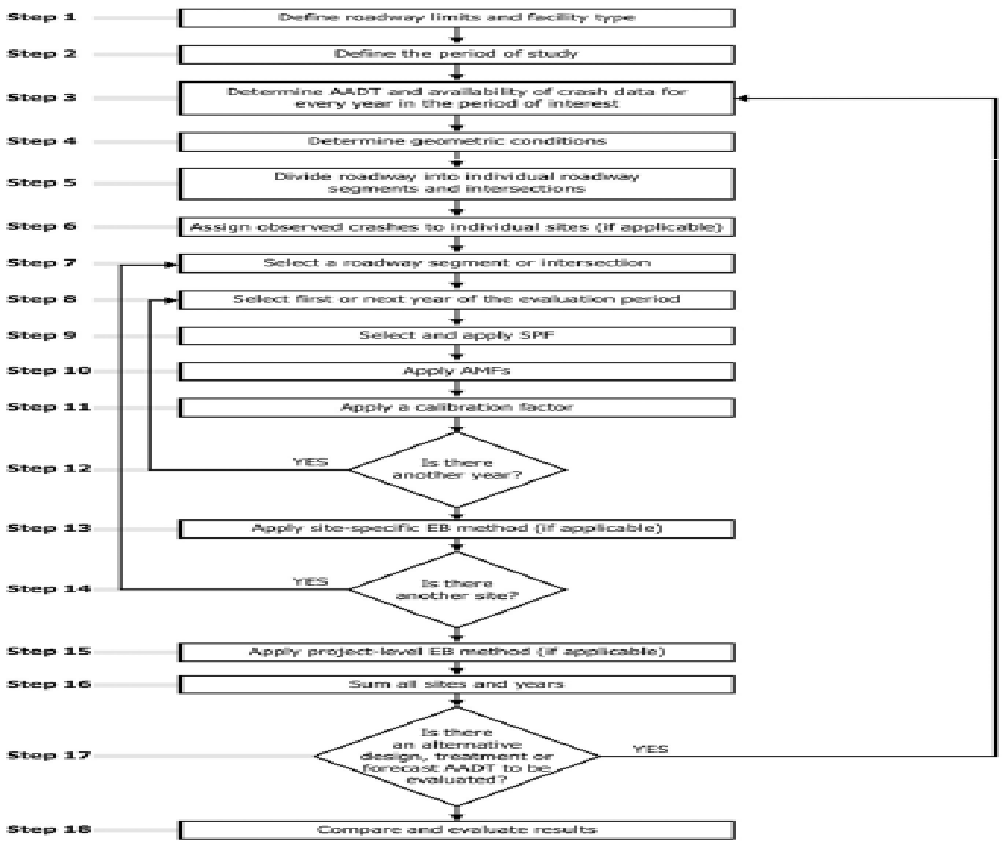
$$N_{spf\ 4SG} = exp[-5.13 + 0.60 \times ln(AADT_{maj}) + 0.20 \times ln(AADT_{min})]$$

  - The base conditions are
    - No Left turn lanes on any approach
    - No Right turn lanes on any approach

1 ADT_{maj}	2 ADT_{min}	3 SPF Prediction	4 Intersection Approaches with Left-Turn Lanes	5 AMF_1	6 Intersection Approaches With Right-Turn Lane	7 AMF_2	8 Years of Data	9 Predicted Average Crash Frequency	10 Observed Crash Frequency
4000	2000	2.152	1	0.67	1	0.98	3	4.240	4
3000	1500	1.710	0	1.00	2	0.95	2	3.249	5
5000	3400	2.736	0	1.00	2	0.95	3	7.799	10
6500	3000	3.124	0	1.00	2	0.95	3	8.902	5
3600	2300	2.078	1	0.67	1	0.98	3	4.093	2
4600	4500	2.753	0	1.00	2	0.95	3	7.846	8
5700	3300	2.943	1	0.67	1	0.98	3	5.796	5
6800	1500	2.794	1	0.67	1	0.98	2	3.669	4
						Sum		45.594	43
						Calibration Factor (C_i)			0.943

Predictive Method Procedures

- In HSM, the predictive method provides an 18 step procedure to estimate the expected average crash frequency of a site, facility, or roadway network.



Predictive Method Procedures

- Step 1 Determine data needs
  - Facility type
  - Study period
  - Site conditions (geometry, traffic control, etc.)
  - Traffic volume (vehicles/day)
- Step 2 Divide locations into homogeneous segments or intersections
  - Number of lanes
  - Type of intersections
  - Alignment change
  - Change in roadside conditions
  - Change in traffic volume

Predictive Method Procedures

- Step 3 Identify and apply the appropriate SPF

2-Lane Rural Highway

Multilane Rural Highway

Urban Arterial

- Step 4 Apply CMFs to calculated SPF values
  - Review applicable SPF “base case”
  - Determine how study site differs from “base case”
  - Select CMFs for road type and typical features
  - Multiply SPF value by applicable CMFs

Predictive Method Procedures

- Step 5 Apply local calibration factor
  - "C" adjusts SPF-derived crash estimates to reflect local conditions
  - Each SPF requires its unique "C"

Predictive Method Procedures

- Supplemental steps
  - Repeat basic steps (time period)
  - Apply site-specific EB method
  - Repeat basic steps (next study site)
  - Apply project-level EB method
  - Calculate total expected crashes
  - Evaluate alternate design
  - Evaluate and compare results

Summary

$$N_{predicted} =$$

$$SPF \times (CMF_1 \times CMF_2 \times \ldots.) \times C$$

where:

SPF = Safety Performance Function

CMF = Crash Modification Factor

C = Calibration Factor

$$N_{expected} = N_{predicted} \oplus N_{observed}$$